Received: 26 May 2022 | Revised: 9 February 2023 | Accepted: 9 June 2023

DOI: 10.1002/for.3008

RESEARCH ARTICLE

WILEY

Variable selection for classification and forecasting of the family firm's socioemotional wealth

Susana Álvarez-Díez[1] | **J. Samuel Baixauli-Soler**[2] | **María Belda-Ruiz**[2] | **Gregorio Sánchez-Marín**[3]

[1]Department of Quantitative Methods for Economics and Business, Faculty of Economics and Business, University of Murcia, Campus de Espinardo, Murcia, Spain

[2]Department of Management and Finance, Faculty of Economics and Business, University of Murcia, Campus de Espinardo, Murcia, Spain

[3]Department of Economics and Business, Faculty of Economics, Business and Tourism, University of Alcala, Alcala de Henares, Spain

Correspondence
María Belda-Ruiz, Department of Management and Finance, Faculty of Economics and Business, University of Murcia, Campus de Espinardo, 30100 Murcia, Spain.
Email: maria.belda@um.es

Funding information
Fundación Cajamurcia

Abstract

Socioemotional wealth (SEW) refers to those family-centered goals that are likely to have a major influence on the strategic decision-making process and performance of family firms. Many studies have used indirect indicators related to family involvement in ownership and management to measure SEW; meanwhile, others have developed scales to directly measure the level and importance of SEW in family firms. Limitations of both indirect and direct measures of SEW lead empirical research on SEW to be under threat. In the current study, we use random forests to identify the important indicators related to financial and economic decisions, as well as family-related measures, for explaining the family firms' SEW and to design a good prediction model using the smallest set of nonredundant indicators. Our results show that the model that exhibits the minimum out-of-bag sample (OOB) error rate includes variables that refer to the presence of family members in the firm's management positions, long-term nonfinancial debt, personnel expenditures, long-term financial investments, short-term financial debt, average storage period, and accounts receivables. For prediction, the model with a reasonably low estimated classification error includes only three variables, which refer to the presence of family members in the firm's management positions, long-term nonfinancial debt, and accounts receivables.

KEYWORDS

estimated classification error, family firms, financial database, random forests, socioemotional wealth

1 | INTRODUCTION

One of the most relevant and growing areas in family firms concerns financial decisions (Michiels & Molly, 2017; Schickinger et al., 2022). Most previous studies have focused on the difference between family and nonfamily firms providing mixed results regarding debt levels (Ampenberger et al., 2013; López-Delgado & Diéguez-Soto, 2018; Mishra & McConaughy, 1999; Schmid, 2013; Setia-Atmaja, 2010), dividend policy (Attig et al., 2016; Gugler, 2003; Pindado et al., 2012; Setia-Atmaja et al., 2009), and

external equity financing (Poutziouris, 2001; Wu et al., 2007).

Moving away from such traditional distinction between family and nonfamily business, studies have taken a step further toward the analysis of family involvement in ownership and management and financial and economic decisions within the unique context of family firms (Baek et al., 2016; Belda-Ruiz et al., 2022; Blanco-Mazagatos et al., 2007; Jansen et al., 2022; Koropp et al., 2014). Since the introduction of the socioemotional wealth (SEW) theory developed by Gomez-Mejia et al. (2007), the literature has started to identify family involvement with the importance attached to preserving SEW and its effects on decision making (Molly et al., 2019; Schickinger et al., 2022; Vandemaele & Vancauteren, 2015). As the desire to preserve SEW increases, family firms are likely to be more inclined to take less risky strategic decisions to avoid putting their family wealth or SEW at risk (Berrone et al., 2012; Gomez-Mejia et al., 2011). In terms of performance, although the traditional SEW theoretical framework supports that family firms may take decisions that separate from the rational financial logic (Gomez-Mejia et al., 2007), recent studies suggest that the pursuit of family SEW, on average, does not occur at the expense of financial utility and show a positive relationship between SEW and firm performance (Davila et al., 2022).

The fact of studying the SEW framework in family firms and its influence on corporate strategic decisions have led many studies to proxy the SEW construct with indirect measures including, among others, CEO family status (Naldi et al., 2013), generational stage (Vandemaele & Vancauteren, 2015), binary measures for the presence of external managers (Stockmans et al., 2010), or continuous variables capturing the percentage of family members among the firm's managers (Sciascia et al., 2014). But these studies do not directly measure SEW, which supposes a clear shortcoming and a mismatch between SEW as a theoretical construct and its empirical correlate (Schulze & Kellermanns, 2015). There are few attempts to directly assess the SEW construct itself that implies the development of scales and survey data (Berrone et al., 2012; Debicki et al., 2016; Gerken et al., 2022; Hauck et al., 2016; Reina et al., 2022), which are the typical limitations for research of survey data (Prügl, 2019).

The aim of this study is to face the empirical challenge of classifying family firms in family firms with high level and low level of SEW using variable selection given a great variety of indicators related to financial and economic decisions and family-related measures from financial database. We attempt to do that using random forests. Doing that, we contribute to the literature by overcoming the limitations and shortcomings of classical direct and indirect measures in terms of measuring the SEW level (Hauck et al., 2016; Kotlar et al., 2018; Prügl, 2019). Because the evidence shows that SEW preservation in family firms is associated with improving firm performance (Davila et al., 2022), and problems related to both indirect and direct measures of SEW lead empirical research on SEW to be under threat, it is necessary to contribute to the literature by forecasting SEW levels using objective indicators that could reflect the different behavior of family firms with low and high levels of SEW.

The remainder of the paper is structured as follows. In Section 2, we review the literature of SEW and its direct and indirect measures. Then, Section 3 presents the variables and methodology, and the results are included in Section 4. Finally, in Section 5, we formulate the main conclusions of the paper.

2 | LITERATURE REVIEW

2.1 | SEW in family firms

SEW or affective endowments were introduced by Gomez-Mejia et al. (2007) and are defined as "non-financial aspects of the firm that meet the family's affective needs, such as identity, the ability to exercise family influence, and the perpetuation of the family dynasty" (Gomez-Mejia et al., 2007; p. 106). Since then, researchers have focused on both the theoretical and the empirical points of view of the desire of family firms to preserve and pursue SEW (Brigham & Payne, 2019; Chua et al., 2015; Kotlar et al., 2018; Martin et al., 2016; Schulze & Kellermanns, 2015).

Based on behavioral agency model (BAM) (Wiseman & Gomez-Mejia, 1998), SEW framework argues that a reference point in the decision making among family firms is the potential balance for gains or losses in the stock of affective endowment (Gomez-Mejia et al., 2007). Family owners are thus driven to preserve and enhance their family wealth, apart from any financial benefits. SEW preservation incentivizes family firms to avoid risky decisions (such as internationalization, diversification, and R&D investments), even if these decisions are favorable from a financial logic (Martin & Gómez-Mejía, 2016).

The literature points out two opposite explanations of how SEW may affect firm performance. On the one hand, family firms' preferences to preserve SEW may lead decisions that sacrifice financial returns. The preservation of business under family control can lead family owners to restrict external resources and skills that maximize financial wealth (Alessandri et al., 2018). In addition, SEW preservation may also induce family members to use their power to divert business resources to benefit the family

(Schulze et al., 2003; Schulze & Kellermanns, 2015), with specific manifestations of nepotism, cronyism, entrenchment (Kets de Vries, 1993; Villalonga et al., 2015) that negatively influence firm performance.

On the other hand, several studies confirm a positive influence of SEW on firm performance. General desire to preserve SEW among family owners induces them to facilitate transitions to the next generations adopting thus on a long-term orientation (Zellweger et al., 2012) that gives to the accumulation of patient capital and long-term returns in order to achieve a sustainable competitive advantage (Arregle et al., 2007). In addition, SEW can foster high-performance human resource practices among family and nonfamily employees (Peláez-León & Sánchez-Marín, 2022) and by treating nonfamily employees as part of a family and promoting their identification with the organization's culture and values (König et al., 2013). Moreover, SEW preservation pushes family owners to maintain the firm's reputation and project a positive image to the community, providing the necessary organizational status and legitimacy to improve performance (Berrone et al., 2022).

Helping to build consensus about the SEW–performance relationship in family firms, the recent meta-analysis of Davila et al. (2022), covering 350 studies during 2007–2020 and 2,959,720 firm-year observations, finds that SEW has a generally positive relationship with firm performance, denying that the preservation of family wealth in family firms jeopardize financial results. In addition, analyzing some key contingent variables in this relationship, these authors show that family firms' risk aversion to preserve SEW still serves as an inducement to outperform competitors and that idiosyncratic corporate governance practices that characterized family firms—and that usually have been identified as inefficient—tend to have a positive effect on firm performance, further improving the welfare of stakeholders. Finally, this meta-analysis shows that most SEW subdimensions (except for dynastic succession) have also a positive relationship with firm performance.

Then, noneconomic goals captured by the SEW construct are the pivotal frame of reference in the family firms' strategic decisions (Gomez-Mejia et al., 2014; Naldi et al., 2013; Patel & Chrisman, 2014), with the subsequent positive effect on firm performance (Davila et al., 2022).

2.2 | Indirect proxies for SEW preservation in family firms

Despite the increasing popularity and generality of the theoretical perspective of SEW and the empirical evidence on their effects on family firms' decisions and performance, literature has remarked problems and challenges about measuring SEW level properly (Berrone et al., 2012; Debicki et al., 2016; Miller & Le Breton-Miller, 2014; Schulze & Kellermanns, 2015; Lardon et al., 2017). The exact construct of interest, its unidimensional or multidimensional nature, and its measurement is still an unresolved question (Brigham & Payne, 2019; Reina et al., 2022). Regarding this last challenge, prior literature has used indirect measurements of SEW related to several variables of family involvement in ownership and management. Some of the most used variables are the following:

2.2.1 | Family ownership

A certain percentage of family ownership is usually used in the literature to consider a firm as a family firm, a percentage that ranges from 5% in publicly traded firms (Berrone et al., 2010; Chrisman & Patel, 2012; Gomez-Mejia et al., 2018; Patel & Chrisman, 2014), to 20–25% (Cruz et al., 2014; Deslandes et al., 2016; Kotlar et al., 2018; Leitterstorf & Rau, 2014; Setia-Atmaja, 2010) or 50% in private firms (Michiels et al., 2015, 2017). As family ownership increases, SEW preservation and its effects on the firm's decisions will be strongly present in the day to day of the firm. Using this measure, a family firm should be considered with high level of SEW if the family ownership is high and low level of SEW otherwise.

2.2.2 | CEO family status

Having a family member in the CEO position of the firm is considered one of the main SEW-preserving mechanisms (Naldi et al., 2013). Compared with nonfamily CEOs, family CEOs want to protect their SEW and keep control over the firm and therefore are more concerned with the protection of family-centered goals (Baixauli-Soler et al., 2021; Huybrechts et al., 2013). Using this measure, a family firm should be considered with high level of SEW if the CEO is a family member and low level of SEW otherwise. CEO family status, through a dummy variable, has been included in the literature as proxy for SEW to analyze its effects on environmental issues (Berrone et al., 2010), financing and dividend policy (Lardon et al., 2017; Vandemaele & Vancauteren, 2015), firm performance (Naldi et al., 2013), or earnings management (Stockmans et al., 2010), among others.

2.2.3 | Family members in other management positions

The value attached to the firm's SEW also differs according to the presence of family members in other management

positions beyond the CEO. Similarly, when there are more family members in the top management team of the firm or on the board of directors, their desire to preserve SEW will be higher and will affect strategic choices and performance (Gomez-Mejia et al., 2007). Using this measure, a family firm should be considered with high level of SEW if the number of family members in management positions is high and low level of SEW otherwise. In this line, Stockmans et al. (2010) consider the presence of external managers through a dummy variable in their study about earning managements, and other researchers consider, to distinguish a family firm from a nonfamily firm, the presence of family members on the board of directors (Cruz et al., 2014; Gomez-Mejia et al., 2010).

2.2.4 | Generational stage

According to Gomez-Mejia et al. (2011), the family stage that oversees the firm is considered an important contingency variable affecting the desire to preserve SEW in the firm and managerial decisions and performance. SEW preservation increases in importance when the firm is managed by the first generation; it is when the identification with the firm is stronger and noneconomic goals are the unique point of reference in managerial choices (Gomez-Mejia et al., 2011). In later generations, economic and financial goals may become more important than SEW preservation (Sciascia et al., 2014). Using this measure, a family firm should be considered with high level of SEW if the firm is managed by the first generation and low level of SEW otherwise. Many previous SEW-related studies have considered in their empirical analysis dummy variables that capture the family generation managing the firm (Gomez-Mejia et al., 2007; Sciascia et al., 2014; Stockmans et al., 2010; Vandemaele & Vancauteren, 2015).

All the above set of family-related variables attempt to indirectly measure SEW preservation in family firms to analyze, in different contexts, potential effects on financial and economic decisions and performance. The difficulties of directly assessing SEW preservation lead previous studies to rely on such indirect variables, but the literature criticizes the use of indirect measures for SEW (Miller & le Breton–Miller, 2014; Schulze & Kellermanns, 2015). As Hauck et al. (2016) indicate, the main limitations of these indirect family variables are the following: Family firms with the same management and ownership structure may have different levels of affective endowments; SEW preservation may not be positive and proportionally linked to variables of family ownership or management, being a multidimensional, and no lineal relationship.

2.3 | Direct measurement scales of SEW preservation in family firms

To overcome the limitations of the use of indirect proxies for the operationalization of SEW, a few previous studies have developed scales to directly measure the level and importance of SEW in family firms (for a review, see Reina et al., 2022). The first scale was developed by Berrone et al. (2012), known as the FIBER scale, who proposed 27 items to capture the family control and influence, the identification and emotional attachment of family members with the firm, binding social ties (including relationships with other institutions or suppliers and social activities of the considerations of nonfamily employees as part of the family), and renewal of family bonds through next generations (Berrone et al., 2012). Using cross-sectional quantitative survey data in the German-speaking area, Hauck et al. (2016) empirically validate the FIBER scale and offer a shorter one, called the REI scale, which only includes nine items of the initial set proposed by Berrone et al. (2012). Recently, Gerken et al. (2022) have replicated and extended the first validation of the FIBER scale undertaken by Hauck et al. (2016) offering an improved (short) SEW scale. Because the initial FIBER scale includes too many items with the limitations of having a nonpractical length, other studies consider a selection of the items included in the FIBER scale to directly measure SEW (Kallmuenzer et al., 2018).

Another important scale in the literature is developed by Debicki et al. (2016). These researchers propose and validate their own scale to measure the importance of SEW in family firms, which has been used in later studies. They attempt to capture relevant noneconomic aspects such as the importance of how the family is perceived by the community, the importance of preserving the firm's control in the hands of family members, and the desire to fulfill a broader range of obligations toward family members to enhance the harmony of the family at large rather than just the members directly involved with the business.

Other studies have also developed and used in their studies one-dimensional scales to measure SEW. It is the case of Goel et al. (2013) who measure SEW through four items related to maintaining the firm family traditions and character, providing or maintaining jobs in the firm from the family members, independence in ownership, and independence in management. Although the first two questions represent the perpetuation of the family dynasty, the other two proxy for the ability to exercise family influence and maintain family control. Later, the scale of Goel et al. (2013) has been used by other previous studies in different contexts of family firms, including the entrepreneurial orientation–performance relationship or

the appointment of nonfamily managers (Schepers et al., 2014; Vandekerkhof et al., 2015). Based on the literature, Cabrera-Suárez et al. (2014) also consider five items that capture the dimensions of Berrone et al. (2012) to analyze factors driving nonfinancial goals in family firms, and more recently, Molly et al. (2019) examine financial decisions in family firms by measuring SEW with four questions related to the success of the business transfer to the next generation, preservation of family control and independence, minimization of family conflicts, and provision of work to family members.

Despite the existence of the above attempts of developing direct measurement scales of SEW to overcome the problems related to the use of indirect measures of SEW, limitations also emanate from these scales, including problems related to the length of the scale, a low response rate or the impossibility to implement these SEW measurement retrospectives at the time of other available data to conduct our research (Kotlar et al., 2018; Prügl, 2019). Thus, problems related to both indirect and direct measures of SEW lead empirical research on SEW to be under threat.

3 | METHODOLOGY

3.1 | Random forests

Given a large set of indicators, in this section, we describe the procedure to identify the most relevant indicators, even with redundancy, related to financial and economic decisions and family-related measures available in financial databases for explaining the SEW level in family firms and to design a good prediction model using the smallest set of nonredundant indicators that achieve good predictive performance.

More specifically, we face a classification problem, whether a family firm has a high/low level of SEW, implementing random forest models (Breiman, 2001, 2004), which rely on decision trees to produce one optimal predictive model. Nowadays, decision tree-based methods (Breiman et al., 1984) are widely used to generate predictions because of their good results when multiple features interact between them in a complex and nonlinear manner (Banfield et al., 2007). Many empirical studies have been conducted using random forests in recent years (Chen et al., 2021; Díaz-Uriarte & Alvarezde Andrés, 2006; Gao et al., 2019; Tabatabaee Malazi & Davari, 2018; Uddin et al., 2020; Wang et al., 2018; among others).

We focus on random forests mainly because they work well for high-dimensional problems for which the number of variables hugely exceeds the sample size (Genuer et al., 2008) without overfitting. In addition, these models allow us the quantification of the variable importance.

Random forests are a class of important ensemble methods that combines bootstrapping and aggregation to generate an optimal predictive model. Given a training set bootstrap training sets are generated as bagging (Breiman, 1996). In contrast to bagging, a number of features are randomly selected at each pseudosample to grow a tree using CART methodology (Breiman et al., 1984) without pruning, preventing a very influential feature from dominating many trees. Consequently, each tree will be split based on slighted different samples and different features, providing decorrelated trees and producing a more accurate predictor. After each randomized tree has been grown, an algorithm of aggregation is used to get the final predictor. In a problem of classification, it will be the mode of predictions of the subsample decision trees.

As the trees are trained using bootstrapped subsamples, on average, each training uses only about two-thirds of the whole sample. The remaining third is not used for growing a tree, is called out-of-bag sample (OOB), and is used to compute the OOB error to test the performance of the model. Breiman (2001) showed that the OOB error is an unbiased estimation of the prediction error, which measures how accurately the random forest predicts unseen data. Figure 1 shows how the decision trees are in a random forest model.

As Genuer et al. (2010) pointed out, the quantification of the variable importance is relevant to rank the variables and select some of them to estimate the model as well as to determine which variables explain the most response, helping with better understanding of the solved problem. Variable importance may lead to model accuracy improvements by using the variable selection. Biau (2012) showed that the model suggested by Breiman (2004) is consistent and adapts to sparsity, in the sense that its rate of convergence depends only on the number of strong features and not on how many noise variables are. Some articles focused on variable selection in classification problems based on random forests are Guyon et al. (2002), Rakotomanonjy (2003), Díaz-Uriarte and Alvarez de Andrés (2006), Poggi and Tuleau (2006), Genuer et al. (2010), Speiser et al. (2019), and Chavent et al. (2021).

Genuer et al. (2010) provided some insights about the behavior of the variable importance index based on random forests, and to use it, they proposed a two-step algorithm for variable selection starting from a variable importance ranking based on the permutation importance measure. Following Genuer et al. (2010), we will use random forests in two steps: first, to rank each variable's importance to identify the most highly related important variables to the dependent variable (high/low level of SEW) for interpretation purpose, and second, to

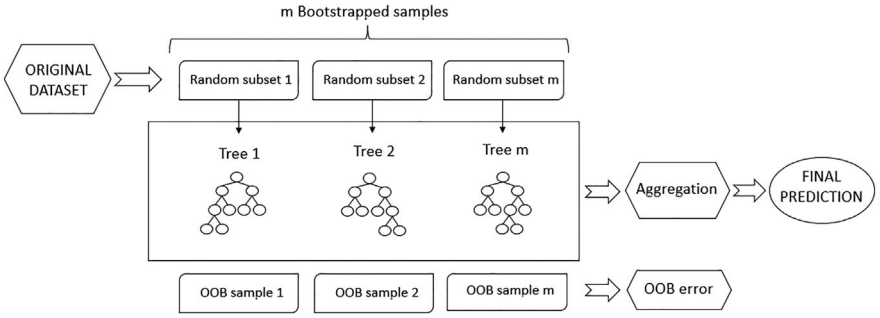

FIGURE 1 Decision tree in a random forest model.

find the smallest set of nonredundant variables (indicators from financial databases) to achieve a good prediction of the dependent variable. The three-step procedure can be described as follows:

- Thresholding step: For each variable I^k, $k=1,...,K$, compute the variable importance, $VI(I^k)$, using the following expression:

$$VI(I^k)=\frac{1}{ntree}\sum_{j=1}^{ntree}\left(\widetilde{errOOB}_j^k-errOOB_j\right)$$

where $ntree$ is the number of trees of the forest, $errOOB_j$, $j=1,...,ntree$ is the misclassification rate for the classification of the tree j on the OOB_j sample, and $\widetilde{errOOB}_j^k$ is the misclassification rate for the classification of the tree j when the values of I^k are randomly permuted in the OOB_j sample.

Rank each variable's importance (averaged over r random forest runs), select the first m variables in descending order of importance, and remove the $K-m$ variables that are less important.

- Interpretation step: Grow a nested sequence of forests, $RF^1,...,RF^m$, from using only the most important variable (RF^1) to use the m variables (RF^m). Compute $errOOB^{RF^1},...,errOOB^{RF^m}$ for r random forest runs and select the variables I^k, $k=1,...,m'$, leading to the smallest mean $errOOB$ (averaged over r random forest runs). They will be used for interpretation purpose.
- Prediction step: The starting point is the set of variables previously selected for interpretation increasingly ordered by degree of importance, $I^k_{ordered}$, $k=1,...,m'$. Grow a sequence of random forests increasingly, adding variables $I^k_{ordered}$, $k=1,...,m'$, one by one. The added variable will be included in the prediction model whenever the decrease of $errOOB$ is significantly larger than the following threshold:

$$\frac{1}{m-m'}\sum_{k=m'}^{m-1}|errOOB(k+1)-errOOB(k)|,$$

where $errOOB(k)$ is the $errOOB$ computed using the k most important variables.

3.2 | SEW data

SEW variable is obtained using the scale of Debicki et al. (2016) for a sample of 367 private medium-sized Spanish family firms for 2016, and the random forests include 528 economic and financial indicators obtained from SABI (Iberian Balance Sheets Analysis System) database. As shown in panel A of Table 1, these indicators are related to indebtedness, investment, personnel expenditures, dividend policy and reserves, cash holding, and other current assets and liabilities. This set of indicators includes 88 economic and financial ratios as follows: the corresponding values for 2016 (R*number*), the variation rate of each ratio between 2015 and 2016 (VR*number*), the growth rate of each ratio over the period 2012–2016 (GR*number*), and industry-relative indicators including the value for 2016 (IR*number*), the 2015–2016 variation rate (IVR*number*), and the 2012–2016 growth rate (IGR*number*). We also include 10 indirect measures of SEW related to family ownership, CEO family status, other family management positions, and family generational stage (see panel B of Table 1).

Thus, our dataset consists of family-related measures and objective financial indicators obtained from financial databases that could reflect the adoption of economic and financial strategic decisions and performance effects that differ between family firms with high and low levels of SEW (Gomez-Mejia et al., 2011). In particular, the following financial and economic dimensions are considered.

3.2.1 | Indebtedness

Previous studies focused on debt decisions in family and nonfamily firms have shown mixed results (Ampenberger et al., 2013; Anderson & Reeb, 2003; Schmid, 2013; Setia-Atmaja et al., 2009). Debt financing

TABLE 1 Economic and financial indicators and indirect proxies for socioemotional wealth (SEW) in family firms.

Panel A: Economic and financial dimensions	
Indebtedness	**Indicators**
Total indebtedness	**R1**: Total debt/Total equity and debt; **R2**: Total equity/Total equity and debt; **R3**: Financial debt/Total equity and debt; **R4**: Financial debt/Total debt; **R5**: Nonfinancial debt/Total equity and debt; **R6**: Nonfinancial debt/Total debt
Long-term indebtedness	**R7**: Long-term financial debt/Total debt; **R8**: Long-term financial debt/Long-term debt; **R9**: Long-term financial debt/Total equity and debt; **R10**: Long-term debt/Total equity and debt; **R11**: Long-term debt/Total debt; **R12**: Long-term nonfinancial debt/Total debt; **R13**: Long-term nonfinancial debt/Total equity and debt; **R14**: Long-term nonfinancial debt/Long-term debt
Short-term indebtedness	**R15**: Short-term financial debt/Total debt; **R16**: Short-term financial debt/Short-term debt; **R17**: Short-term financial debt/Total equity and debt; **R18**: Short-term debt/Total equity and debt; **R19**: Short-term debt/Total debt; **R20**: Short-term nonfinancial debt/Total debt; **R21**: Short-term nonfinancial debt/Total equity and debt; **R22**: Short-term nonfinancial debt/Short-term debt; **R23**: Short-term debt/Long-term debt; **R24**: Short-term financial debt/Long-term financial debt
Financial expenses (and incomes)	**R25:** Financial expenses/Total equity and debt; **R26**: Financial expenses/Total debt; **R27**: Financial expenses/Financial debt; **R28**: Financial expenses/Net sales; **R29**: Financial incomes/Total assets; **R30**: Financial incomes/Net sales; **R31**: (Financial incomes-Financial expenses)/Total assets; **R32**: (Financial incomes-Financial expenses)/Net sales; **R33**: (Financial incomes-Financial expenses)/Earnings before interest and taxes; **R34**: Financial expenses/Earnings before interest and taxes
Investment	**Indicators**
Noncurrent assets	**R35**: Tangible assets/Total assets; **R36**: Tangible assets/Noncurrent assets; **R37**: Noncurrent financial assets/Total assets; **R38**: Noncurrent financial assets/Noncurrent assets; **R39**: Investment property/Total assets; **R40**: Investment property/Noncurrent assets; **R41**: Noncurrent assets/Total assets
Other noncurrent assets	**R42**: Other noncurrent assets/Noncurrent assets; **R43**: Other noncurrent assets/Total assets; **R44**: Total equity/Noncurrent assets
Personnel expenditures	**Indicators**
Personnel expenses	**R45**: Personnel expenses/Net sales; **R46**: Personnel expenses/Other operating expenses; **R47:** Personnel expenses/Number of employees; **R48**: Personnel expenses/Noncurrent assets
Employees	**R49**: Number of employees/Net sales; **R50**: Number of employees/Other operating expenses; **R51**: Number of employees/Noncurrent assets
Dividend policy	**Indicators**
Dividends and reserves	**R52**: Total ordinary dividends/Net income; **R53**: Reserves/Total equity; **R54**: Reserves/Total equity and debt; **R55**: Reserves/Net income; **R56**: Reserves/Noncurrent assets
Liquid assets (cash holdings)	**Indicators**
Cash	**R57:** Cash/Total assets; **R58**: Cash/Current assets; **R59**: Cash/Short-term debt; **R60**: Cash/Net sales
Cash and cash equivalents	**R61:** Cash and cash equivalents/Total assets; **R62**: Cash and cash equivalents/Current assets; **R63**: Cash and cash equivalents/Short-term debt; **R64**: Cash and cash equivalents/Net sales
Other current assets and liabilities	**Indicators**
Current assets	**R65**: Current assets/Total assets; **R66**: Current assets/Short-term debt; **R67**: (Current assets-inventories)/Short-term debt; **R68**: (Current assets-inventories-trade receivables)/Short-term debt; **R69**: (Current assets-Short-term debt)/Net sales; **R70**: Current assets/Noncurrent assets
Inventories	**R71:** Inventories/Total assets; **R72**: Inventories/Current assets; **R73**: Inventories/Short-term debt; **R74**: Inventories/Net sales; **R75**: (Inventories/Supplies)*365; **R76**: (Inventories + trade receivables)/Trade payables
Trade receivables	**R77:** Trade receivables/Total assets; **R78**: Trade receivables/Current assets; **R79**: Trade receivables/Short-term debt; **R80**: Trade receivables/Net sales; **R81**: Trade receivables/Trade payables; **R82**: (Trade receivables/Net sales)*365
Short-term financial assets	**R83:** Short-term financial assets/Total assets; **R84**: Short-term financial assets/Current assets; **R85**: Short-term financial assets/Short-term debt; **R86**: Short-term financial assets/Net sales
Trade payables	**R87:** Trade payables/Net sales; **R88**: (Trade payables/supplies)*365

(Continues)

TABLE 1 (Continued)

Panel B: Indirect proxies for SEW in family firms	
Family dimensions	**Proxies**
Family ownership	**A1**: Percentage of family ownership; **A2**: Percentage of family shareholders
CEO family status	**A3**: CEO family status (dummy variable); **A4**: Family CEO shareholder (dummy variable)
Family management	**A5:** Percentage of family managers; **A6**: Family Chair status (dummy variable); **A7**: Percentage of family members on the board of directors
Generational stage	**A8:** First generation managing the firm (dummy variable); **A9:** Second generation managing the firm (dummy variable); **A10**: Third generation managing the firm (dummy variable)

Note: Panel A includes 88 economic and financial indicators (ratios). On the basis of these 88 indicators, the dataset used in random forests consists of 528 indicators formed as follows: the corresponding value of the ratio for 2016 (R*number*), the variation rate of each ratio between 2015 and 2016 (VR*number*), the growth rate of each ratio over the period 2012–2016 (GR*number*), and industry-relative indicators in terms of the value for 2016 (IR*number*), the 2015–2016 variation rate (IVR*number*), and the 2012–2016 growth rate (IGR *number*). Panel B includes 10 indirect proxies for SEW in family firms (A*number*) used in random forests.

is one of the strategic choices influenced by noneconomic factors (Baixauli-Soler et al., 2021; Gomez-Mejia et al., 2011; Koropp et al., 2014). The desire to preserve SEW may lead family firm to use more debt avoiding, in this way, funds from external equity that are associated with losses in family firm's control and SEW (Gottardo & Moisello, 2019; Romano et al., 2001). The presence of family members in management positions could increase such leverage (Gottardo & Moisello, 2014; Lardon et al., 2017). But a negative effect of SEW on debt is also possible. Passing up growth, if necessary (López-Gracia & Sánchez-Andújar, 2007), family firms may be more inclined to use internal financing and lower levels of debt in order not to face the financial risk associated with debt and creditor monitoring (Schmid, 2013). Family firms are more likely to use lower levels of debt, and this negative effect is stronger in the case of short-term debt and financial debt (Mishra & McConaughy, 1999; Molly et al., 2019). The heterogeneity existing in family firms may be the reason for the lack of concluding remarks in terms of indebtedness (Berrone et al., 2010; Chua et al., 2012; De Massis et al., 2013). Because family firms with high and low SEW levels may exhibit a different financial behavior, our models consider a great variety of variables that capture the different dimensions of debt, including ratios of total debt, short- and long-term debt, financial and nonfinancial debt, and financial expenses.

3.2.2 | Investment (noncurrent assets)

Family firms are usually risk averse and are more inclined to invest in long-term physical assets (Anderson et al., 2012) instead of intangible assets such as R&D (Chrisman & Patel, 2012; Poutziouris, 2001; Sciascia et al., 2015). Contrary to R&D investment, long-term tangible investments allow family members to preserve SEW, assure it for the next generations, and no threats to family's control (Gomez-Mejia et al., 2011, 2014). Then, family firms that present different levels of SEW may show a different behavior regarding the types of investment undertaken by them. We consider several ratios capturing tangible assets, investment property, financial investment, and other noncurrent assets.

3.2.3 | Personnel expenditures

SEW preservation has also a significant effect on human resource management practices (Gomez-Mejia et al., 2011). With respect to employee compensation, the literature reveals that compensation levels (or personnel expenditures) in family firms are lower compared with professionally managed family firms and nonfamily firms (Carrasco-Hernandez & Sánchez-Marín, 2007). Variable compensation has also a smaller importance in compensation packages of employees in family firms (Gomez-Mejia et al., 2003). In the family context, the traditional criterion for setting compensation levels is seniority, which involves considering the loyalty to the firm and the family as factor determining personnel expenses (Sánchez-Marín et al., 2019). Moreover, under the SEW framework, to establish compensation levels, the family firm considers the degree of enhancing the family's harmony, fulfilling family obligations and the family's agenda (Beehr et al., 1997; Gomez-Mejia et al., 2011). Then, family firms in which SEW preservation is more or less important may be heterogeneous in terms of personnel expenses. For this reason, different ratios of human resource policy and personnel expenses have been included in the models.

3.2.4 | Dividend policy

Linked to financial decisions, the literature focused on the distinction between family and nonfamily firms does not provide consistent findings regarding dividend policy (Attig et al., 2016; Gugler, 2003; Setia-Atmaja, 2010). Within the family firm universe, governance practices, family involvement, and the degree of the firm's professionalization influence dividend policy (González et al., 2014; Michiels et al., 2015, 2017). In addition, indirect measures of SEW have been used in the prior literature to examine its effect on dividends. In family firms, dividend payout is lower when the CEO is a member of the family and in the presence of a family-dominated board, being still lower in the case of earlier generational stages (Vandemaele & Vancauteren, 2015). Thus, SEW levels shape dividends in family firms (Belda-Ruiz et al., 2022), and the model of this paper includes several ratios to capture dividend and retained earning policy.

3.2.5 | Liquid assets (cash holdings)

The cash holding policy of family firms is the result not only of financial objectives but also of the firm's nonfinancial objectives. Family firms tend to accumulate more cash than nonfamily firms because of the desire of the first ones to perpetuate the family for future generations (Durán et al., 2016). Steijvers and Niskanen (2013) find that descendant CEOs are associated with higher cash holding than founder CEOs, and this effect is stronger in the case of low ownership dispersion. The literature also shows that cash levels maintained by family firms are higher when the firm faces problems related to family succession or when the founder is involved in the firm's management (Liu et al., 2015). Then, precautionary motives including in the SEW perspective seem to be behind the family firm's incentives to hold more cash (Gomez-Mejia et al., 2007). Several variables of cash and cash equivalents are included in the models to forecast SEW levels in family firms.

3.2.6 | Other current assets and liabilities

A set of indicators of trade credit policy (accounts receivables), accounts payables as source of financing, inventory policy, short-term financial assets, and working capital management has been included in the models to characterize the behavior of family firms with different levels of SEW. Low inventories and the fact of reducing the number of days accounts receivable are associated with increasing profitability (Deloof, 2003), whereas other studies show that profitability increases with investment in receivables. The literature is consistent with a different current financial and asset structure between private family firms and their nonfamily counterparts (Poutziouris, 2001).

4 | RESULTS

The final sample consists of 528 financial and economic indicators and 10 family-related variables as potential factors of the level of SEW for 367 family firms. The response variable is a categorical variable, with two categories, high level of SEW (HIGH) and low level of SEW (LOW), that was created from the underlying continuous SEW variable. We have chosen the 75th percentile as the cutoff point to assign each observation of the SEW variable to a category.

We have used the VSURF (Variable Selection Using Random Forests) package in R (Genuer et al., 2015) to remove irrelevant variables to classify the level (HIGH/LOW) of SEW of family firms, to select the relevant ones for interpretation, and to find the smallest set of variables for prediction. VSURF package uses the random forests permutation-based score of importance to rank the variables.

In the thresholding step, 50 random forests were grown with 2000 trees in each of them. The number of variables randomly sampled in each split was one third of all variables. Figure 2a shows the mean variable importance over the 50 runs for each variable in descending order. The threshold value for variable importance is computed as the minimum prediction value given by a pruned CART tree fitted to the curve of the standard

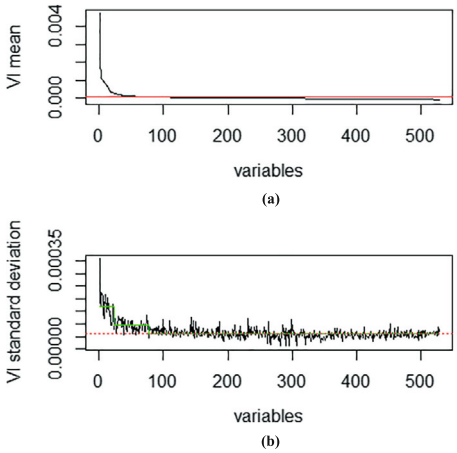

FIGURE 2 (a) Thresholding step. Mean variable importance in descending order. (b) Thresholding step. Variable importance standard deviations, CART tree, and threshold value.

deviations of variable importance. Figure 2b plots the variable importance standard deviations, the CART tree fitted to them (green line), and the threshold value (red dotted line). The relevant variable standard deviations are larger than those of the irrelevant variables, which are close to zero. The threshold is equal to 0.00006073. Consequently, only variables with a mean variable importance larger than 0.00006073 are kept.

Figure 3 zooms in on the thresholding step to improve the visualization.

The number of selected variables at the thresholding step was 102 variables, so 436 irrelevant variables were eliminated from the dataset. This criterium is conservative because it leads to selecting more variables than necessary in order to refine the selection in the interpretation step.

The 102 variables selected in the thresholding step were used in the second step, which is the interpretation step. In this step, nested random forests were grown with 2000 trees each of them, starting with a model that includes only the most relevant variable and ending with one that includes the 102 variables. It was run 25 times to find a sequence of mean OOB errors. The variables selected for interpretation purpose are those corresponding to the model that exhibits a mean OOB error less than the minimum mean OOB error of the sequence of mean OOB errors plus its standard deviation. Figure 4 shows that the OOB error reaches its minimum when the nine first variables are included (see vertical red line) and its value is 0.2131.

The best model for interpretation purpose was provided by keeping the first nine variables, which were A5 (family management), IGR14 (indebtedness), GR14 (indebtedness), R45 (personnel expenditures), IR38 (investment), IVR15 (indebtedness), R75 (other current assets and liabilities), IR75 (other current assets and liabilities), and IR78 (other current assets and liabilities) indexes. The variable A5 is a continuous variable that captures the presence (in percentage) of family members in the firm's management positions. It is measured by the number of managers who belong to the family divided by the total number of the firm's managers. IGR14 captures the 2012–2016 growth rate of long-term debt (maturity of over 1 year), which is nonfinancial (other types of debt not associated with bank loans) with respect to long-term debt and measured in relation to the average of the industry. Therefore, this ratio does not include banking financing. GR14 is computed as the 2012–2016 growth rate of long-term nonfinancial debt divided by long-term debt. In contrast to IGR14, GR14 is not measured in relation to the average of the industry. R45 captures the firm's personnel expenditures with respect to net sales for 2016, that is, personnel expenditures divided by net sales. IR38 is also computed in relation to the average of the industry and captures noncurrent financial assets with respect to the total value of the firm's noncurrent assets, that is, the firm's long-term financial investments for which the full value will not be realized within the accounting year. Said another way, assets whose benefits will be realized over more than 1 year and cannot easily be converted into cash. IVR15 is the bank financing with a maturity of less than 1 year divided by the firm's total debt, measured as 2015–2016 variation rate and in relation to the average of the industry, that is, short-term financial debt divided by total debt. R75 captures the firm's average storage period. In particular, it is measured as $\frac{Inventories}{Supplies} \times 365$, whereas IR75 is the same but an industry-relative indicator. Finally, IR78 is measured as trade (or accounts) receivables divided by current assets, and it is measured in relation to the average of the industry. Accounts receivables are created when a firm lets a buyer purchase their goods or services on credit. Current assets are the firm's assets that are expected to be conveniently sold, consumed, used, or exhausted through standard business operations within 1 year.

The selection was refined in the prediction step to three variables, which were A5, GR14, and IR78. The removed variables did not improve the model enough to be included, and they are correlated with those kept so their information might be already provided by them. The mean OOB errors, averaged over 25 runs, were computed, and Figure 5 shows that the minimum value is reached when those three variables were considered. Its value is equal to 0.2211, and it is the estimated classification error. Hence, the model predicts outcomes reasonably well.

Our results are according to the existing literature. It must be highlighted that the VURF algorithm selected

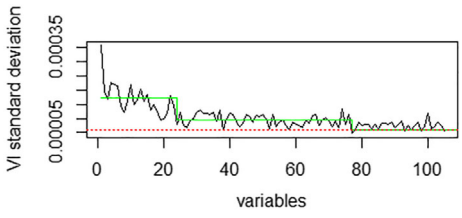

FIGURE 3 Zoom of Figure 2b.

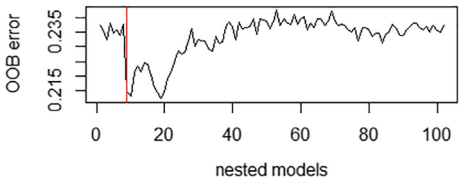

FIGURE 4 Number of selected variables in the interpretation step.

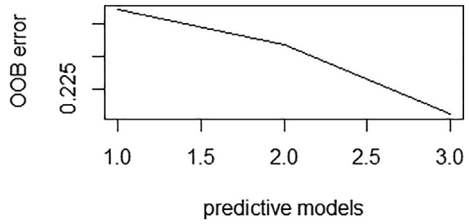

FIGURE 5 Number of selected variables in the prediction step.

the A5 variable, and some empirical studies as Berrone et al. (2012), Huybrechts et al. (2013), Lardon et al. (2017), and Belda-Ruiz et al. (2022), among others, concluded that the family control through family members in top management positions is associated with increasing the level of SEW. It is also interesting to note that related to indebtedness, the literature pointed out that family firms will favor lower levels of financial debt because of control risk considerations in order to preserve SEW (Baixauli-Soler et al., 2021; Molly et al., 2019). In that vein, our results show that GR14 is also relevant to predict the level of SEW. Finally, the importance of the existing difference in trade receivables with respect to total current assets in private family firms with high and low levels of SEW is reflected in the IR78 variable.

To sum up, it should be noted that we have greatly reduced the dimensionality of the classification problem and that the three selected variables in the prediction step allow us to classify a family firm in one of the two categories, being the estimated classification error reasonably low.

To complete the empirical analysis, the three variables extracted by the VSURF algorithm were used for logistic regression to estimate the probability of the response variable (high/low level of SEW) as a function of the three selected variables for prediction. We have generated a logistic model that, based on the variables A5, GR14, and IR78, predicts the probability that a family firm has high level of SEW with respect to the three independent variables. We have selected the 70% of the data for fitting the model (the first 258 observations) and 30% for predicting (testing sample).

The coefficients of the logistic model are estimated according to a maximum likelihood approach for the given values of the variables $A5_i$, $GR14_i$, and $IR78_i$, $i = 1, \ldots, 258$. In Table 2, the results from the estimation of a logit model are presented.

The estimated probability of *high SEW* is

$$\widehat{P(high\,SEW_i)} = \frac{e^{-3.1991+0.01887A3_i+1.2871CR22_i+0.5785NSR76_i}}{1+e^{-3.1991+0.01887A3_i+1.2871CR22_i+0.5785NSR76_i}}$$

The coefficient estimates $\widehat{\beta}_1 = 0.01887$, $\widehat{\beta}_2 = 1.2871$, and $\widehat{\beta}_3 = 0.5785$ are positive, which means that increasing

TABLE 2 Results from estimation of a logit model for the probability of high level of socioemotional wealth (SEW).

	Coefficient	Standard error	*p*-value
A5	0.0188	0.0051	0.0002
GR14	1.2871	0.5353	0.0162
IR78	0.5785	0.2776	0.0371
Constant	−3.1991	0.5722	2.26e-08
Likelihood ratio	24.451		0.00002
Number of firms	258		

Note: *A5* indicates the number of managers who belong to the family divided by the total number of the firm's managers. *GR14* indicates the long-term nonfinancial debt divided by long-term debt (2012–2016 growth rate). *IR78* indicates the trade (or accounts) receivables divided by current assets (in relation to the average of the industry).

A5, GR14, or IR78 will be associated with increasing the probability that a family firm has high level of SEW. The estimated coefficient $\widehat{\beta}_1 = 0.01887$ indicates that one point increase in A5 implies an average change of 0.01887 in the log odds of the response variable, one point increase in GR14 implies an average change of 1.2871 in the log odds of the response variable, and one point increase in IR78 implies an average change of 0.5785 in the log odds of the response variable.

As the *p*-values of A5, GR14, and IR78 are all less than 0.05, at a 5% level, the three independent variables are statistically significant. Moreover, according to the likelihood ratio test ($p = 0.00002$), we have enough evidence to conclude that the given independent variables are jointly significant to predict the probability that a family firm has high level of SEW.

One of the main applications of a logistic regression model is to classify the response variable according to the predictor values. To achieve this classification, we use the testing sample and determine the optimal cutoff point from which the response variable is considered to belong to the level *high level of SEW*. At the optimal cutoff point, both the sensitivity (true positives) and specificity (true negatives) are maximized. In particular, the optimal cutoff point is localized at the value of 0.4044. In other words, an observation is assigned to the group *high level of SEW* if $\widehat{P}(\text{SEW} = \text{high level of SEW}) > 0.4044$ and to the group *low level of SEW* otherwise. Given the optimal cutoff point, we calculate the misclassification rate (0.2018), defined as the percentage mismatch of predicted versus actuals. We got a classification rate of 80%.

5 | CONCLUSION

In this article, an analysis of variable selection and classification of socioemotional wealth data from financial

database is carried out. For a wide dataset consisting of direct and indirect measures of family firms' SEW and financial and economic indicators, we performed variable selection using the VURF method to select the relevant ones for interpretation, computed the OOB error rate, and found the lowest number of variables to classify a family firm in one of the two categories with a low estimated classification error.

Our results show that the model that exhibits the minimum OOB error rate includes variables that refer to the presence of family members in the firm's management positions, long-term nonfinancial debt, personnel expenditures with respect to net sales, long-term financial investments, short-term financial debt, average storage period, and accounts receivables.

For prediction, the model with a reasonably low estimated classification error includes only three variables, which refer to the presence of family members in the firms' management positions, the growth rate of long-term nonfinancial debt divided by long-term debt, and trade (or accounts) receivables with respect to total current assets (in relation to the average of the industry). Then, we have performed a classification problem, reducing drastically the dimensionality. Also, we have checked the predictive ability of the three selected predictors using a logistic regression model, and as we expected, the classification rate was reasonably high. From the logistic regression analysis, it can be concluded that the three variables selected for prediction are directly related to the dependent variable and individually and jointly significant to predict the probability that a family firm has high level of SEW.

Thus, it can be concluded that the behaviors that reflect in a better way high level of SEW preservation in family firms, with its subsequent effect of firm performance, are related to family members exerting control over the firm's strategic decisions through occupying executive positions, avoid financial debt in the long term and therefore financial risk associated with risk and creditor monitoring, and let buyers purchase the goods or services of the firm on credit with better terms with respect to firms that belong to the same industry. Because of the positive SEW–performance relationship shown in previous studies, those three variables could be, therefore, key aspects in the performance of family firms.

ACKNOWLEDGMENTS

Financial support from the Fundación Cajamurcia (Spain) is acknowledged.

DATA AVAILABILITY STATEMENT

The data that support the findings will be available in Zenodo at https://zenodo.org/badge/DOI/10.5281/zenodo.7624551.svg following an embargo from the date of publication to allow for commercialization of research findings.

ORCID

Susana Álvarez-Díez https://orcid.org/0000-0003-3013-0116
J. Samuel Baixauli-Soler https://orcid.org/0000-0003-2444-6894
María Belda-Ruiz https://orcid.org/0000-0002-1277-923X
Gregorio Sánchez-Marín https://orcid.org/0000-0003-0175-814X

REFERENCES

Alessandri, T. M., Cerrato, D., & Eddleston, K. A. (2018). The mixed gamble of internationalization in family and non-family firms: The moderating role of organizational slack. *Global Strategy Journal*, *8*(1), 46–72. https://doi.org/10.1002/gsj.1201

Ampenberger, M., Schmid, T., Achleitner, A., & Kaserer, C. (2013). Capital structure decisions in family firms: Empirical evidence from a bank-based economy. *Review of Managerial Science*, *7*, 247–275. https://doi.org/10.1007/s11846-011-0077-2

Anderson, R. C., Duru, A., & Reeb, D. M. (2012). Investment policy in family controlled firms. *Journal of Banking & Finance*, *36*(6), 1744–1758. https://doi.org/10.1016/j.jbankfin.2012.01.018

Anderson, R. C., & Reeb, D. M. (2003). Founding-family ownership, corporate diversification, and firm leverage. *The Journal of Law and Economics*, *46*(2), 653–684. https://doi.org/10.1086/377115

Arregle, J. L., Hitt, M. A., Sirmon, D. G., & Very, P. (2007). The development of organizational social capital: Attributes of family firms. *Journal of Management Studies*, *44*(1), 73–95. https://doi.org/10.1111/j.1467-6486.2007.00665.x

Attig, N., Boubakri, N., El Ghould, S., & Guedhami, O. (2016). The global financial crisis, family control, and dividend policy. *Financial Management*, *45*(2), 291–313. https://doi.org/10.1111/fima.12115

Baek, H. Y., Cho, D. D., & Fazio, P. L. (2016). Family ownership, control and corporate capital structure: An examination of small capitalization public firms. *Journal of Family Business Management*, *6*(2), 169–185. https://doi.org/10.1108/JFBM-02-2015-0006

Baixauli-Soler, J. S., Belda-Ruiz, M., & Sánchez-Marín, G. (2021). Socioemotional wealth and financial decisions in private family SMEs. *Journal of Business Research*, *123*, 657–668. https://doi.org/10.1016/j.jbusres.2020.10.022

Banfield, R. E., Hall, L. O., Bowyer, K. W., & Kegelmeyer, W. P. (2007). A comparison of decision tree ensemble creation techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(1), 173–180. https://doi.org/10.1109/tpami.2007.250609

Beehr, T. A., Drexler, J. A., & Faulkner, S. (1997). Working in small family businesses: Empirical comparisons to non-family businesses. *Journal of Organizational Behaviour*, *18*(3), 297–313. https://doi.org/10.1002/(SICI)1099-1379(199705)18:3<297::AID-JOB805>3.0.CO;2-D

Belda-Ruiz, M., Sánchez-Marín, G., & Baixauli-Soler, J. S. (2022). Influence of family-centered goals on dividend policy in family firms: A socioemotional wealth approach. *International Entrepreneurship and Management Journal*, *18*(4), 1503–1526. https://doi.org/10.1007/s11365-021-00741-x

Berrone, P., Cruz, C., & Gomez-Mejia, L. R. (2012). Socioemotinal wealth in family firms: Theoretical dimensions, assessment

approaches, and agenda for future research. *Family Business Review*, *25*(3), 258–279. https://doi.org/10.1177/0894486511435355

Berrone, P., Cruz, C., Gomez-Mejia, L. R., & Larraza-Quintana, M. (2010). Socioemotional wealth and corporate responses to institutional pressures: Do family-controlled firms pollute less? *Administrative Science Quarterly*, *55*, 82–113. https://doi.org/10.2189/asqu.2010.55.1.82

Berrone, P., Duran, P., Gómez-Mejía, L., Heugens, P. P., Kostova, T., & van Essen, M. (2022). Impact of informal institutions on the prevalence, strategy, and performance of family firms: A meta-analysis. *Journal of International Business Studies*, *53*, 1153–1177. https://doi.org/10.1057/s41267-020-00362-6

Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, *13*(1), 1063–1095.

Blanco-Mazagatos, V., de Quevedo-Puente, E., & Castrillo, L. A. (2007). The trade-off between financial resources and agency costs in the family business: An exploratory study. *Family Business Review*, *20*(3), 199–213. https://doi.org/10.1111/j.1741-6248.2007.00095.x

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L. (2001). Random forests. *Machine Learning*, *24*, 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2004). Population theory for boosting ensembles. *The Annals of Statistics*, *32*(1), 1–11. https://doi.org/10.1214/aos/1079120126

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. New York: Chapman & Hall.

Brigham, K. H., & Payne, G. T. (2019). Socioemotional wealth (SEW): Questions on construct validity. *Family Business Review*, *32*(4), 326–329. https://doi.org/10.1177/0894486519889402

Cabrera-Suárez, M. K., Déniz-Déniz, M. C., & Martín-Santana, J. D. (2014). The setting of non-financial goals in the family firm: The influence of family climate and identification. *Journal of Family Business Strategy*, *5*(3), 289–299. https://doi.org/10.1016/j.jfbs.2014.05.003

Carrasco-Hernandez, A., & Sánchez-Marín, G. (2007). The determinants of employee compensation in family firms: Empirical evidence. *Family Business Review*, *20*(3), 215–228. https://doi.org/10.1111/j.1741-6248.2007.00096.x

Chavent, M., Genuer, R., & Saracco, J. (2021). Combining clustering of variable and feature selection using random forests. *Communications in Statistics-Simulation and Computation*, *50*(2), 426–446. https://doi.org/10.1080/03610918.2018.1563145

Chen, Y., Zheng, W., Li, W., & Huang, Y. (2021). Large group activity security risk assessment and risk early warning based on random forest algorithm. *Pattern Recognition Letters*, *144*, 1–5. https://doi.org/10.1016/j.patrec.2021.01.008

Chrisman, J. J., & Patel, P. C. (2012). Variations in R&D investments of family and nonfamily firms: Behavioral agency and myopic loss aversion perspectives. *Academy of Management Journal*, *55*(4), 976–997. https://doi.org/10.5465/amj.2011.0211

Chua, J. H., Chrisman, J. J., & De Massis, A. (2015). A closer look at socioemotional wealth: Its flows, stocks, and prospects for moving forward. *Entrepreneurship Theory and Practice*, *39*(2), 173–182. https://doi.org/10.1111/etap.12155

Chua, J. H., Chrisman, J. J., Steier, L. P., & Rau, S. B. (2012). Sources of heterogeneity in family firms: An introduction. *Entrepreneurship Theory and Practice*, *36*(6), 1103–1113. https://doi.org/10.1111/j.1540-6520.2012.00540.x

Cruz, C., Larraza–Kintana, M., Garcés–Galdeano, L., & Berrone, P. (2014). Are family firms really more socially responsible? *Entrepreneurship Theory and Practice*, *38*(6), 1295–1316. https://doi.org/10.1111/etap.12125

Davila, J., Duran, P., Gómez-Mejía, L. R., & Sánchez-Bueno, M. J. (2022). Socioemotional wealth and family firm performance: A meta-analytic integration. *Journal of Family Business Strategy*. In press, *14*, 100536. https://doi.org/10.1016/j.jfbs.2022.100536

De Massis, A., Frattini, F., & Lichtenthaler, U. (2013). Research on technological innovation in family firms: Present debates and future directions. *Family Business Review*, *26*(1), 10–31. https://doi.org/10.1177/0894486512466258

Debicki, B. J., Kellermanns, F. W., Chrisman, J. J., & Pearson, A. W. (2016). Development of a socioemotional wealth importance (SEWi) scale for family firm research. *Journal of Family Business Strategy*, *7*(1), 47–57. https://doi.org/10.1016/j.jfbs.2016.01.002

Deloof, M. (2003). Does working capital management affect profitability of Belgian firms? *Journal of Business Finance & Accounting*, *30*(3–4), 573–588. https://doi.org/10.1111/1468-5957.00008

Deslandes, M., Fortin, A., & Landry, S. (2016). Payout differences between family and nonfamily listed firms. A socioemotional wealth perspective. *Journal of Family Business Management*, *6*(1), 46–63. https://doi.org/10.1108/JFBM-05-2015-0020

Díaz-Uriarte, R., & Alvarez de Andrés, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, *7*(3), 1471–2105. https://doi.org/10.1186/1471-2105-7-3

Durán, R. F., Lozano, M. B., & Yaman, S. (2016). Is family control relevant for corporate cash holding policy? *Journal of Business Finance & Accounting*, *43*(9–10), 1325–1360. https://doi.org/10.1111/jbfa.12222

Gao, X., Wen, J., & Zhang, C. (2019). An improved random forest algorithm for predicting employee turnover. *Mathematical Problems in Engineering*, *2019*, 1–12. https://doi.org/10.1155/2019/4140707

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2008). Random forests: Some methodological insights. *arXiv: 0811.3619*. https://doi.org/10.48550/arXiv.0811.3619

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Patter Recognition Letters*, *31*(14), 2225–2236. https://doi.org/10.1016/j.patrec.2010.03.014

Genuer, R., Poggi, J. M., & Tuleau–Malot, C. (2015). VSURF: An R package for variable selection using random forests. *The R Journal*, *7*(2), 19–33.

Gerken, M., Hülsbeck, M., Ostermann, T., & Hack, A. (2022). Validating the FIBER scale to measure family firm heterogeneity—A replication study with extensions. *Journal of Family Business Strategy*, *13*, 100497. https://doi.org/10.1016/j.jfbs.2022.100497

Goel, S., Voordeckers, W., Van Gils, A., & van den Heuvel, J. (2013). CEO's empathy and salience of socioemotional wealth in family SMEs—The moderating role of external directors. *Entrepreneurship & Regional Development*, *25*(3–4), 111–134. https://doi.org/10.1080/08985626.2012.710262

Gomez-Mejia, L. R., Campbell, J. T., Martin, G., Hoskisson, R. E., Makri, M., & Sirmon, D. G. (2014). Socioemotional wealth as a

mixed gamble: Revisiting family firm R&D investments with the behavioral agency model. *Entrepreneurship Theory and Practice*, *38*(6), 1351–1374. https://doi.org/10.1111/etap.12083

Gomez-Mejia, L. R., Cruz, C., Berrone, P., & De Castro, J. (2011). The bind that ties: Socioemotional wealth preservation in family firms. *The Academy of Management Annual*, *5*(1), 653–707. https://doi.org/10.5465/19416520.2011.593320

Gomez-Mejia, L. R., Haynes, K., Núñez-Nickel, M., Jacobson, K., & Moyano-Fuentes, J. (2007). Socioemotional wealth and business risks in family-controlled firms: Evidence from Spanish olive oil mills. *Administrative Science Quarterly*, *52*(1), 106–137. https://doi.org/10.2189/asqu.52.1.106

Gomez-Mejia, L. R., Larraza-Kintana, M., & Makri, M. (2003). The determinants of executive compensation in family-controlled public corporations. *Academy of Management Journal*, *46*(2), 226–237. https://doi.org/10.5465/30040616

Gomez-Mejia, L. R., Makri, M., & Kintana, M. L. (2010). Diversification decisions in family-controlled firms. *Journal of Management Studies*, *47*(2), 223–252. https://doi.org/10.1111/j.1467-6486.2009.00889.x

Gomez-Mejia, L. R., Patel, P. C., & Zellweger, T. M. (2018). In the horns of the dilemma: Socioemotional wealth, financial wealth, and acquisitions in family firms. *Journal of Management*, *44*(4), 1369–1397. https://doi.org/10.1177/0149206315614375

González, M., Guzmán, A., Pombo, C., & Trujillo, M. A. (2014). Family involvement and dividend policy in closely held firms. *Family Business Review*, *27*(4), 365–385. https://doi.org/10.1177/0894486514538448

Gottardo, P., & Moisello, A. M. (2014). The capital structure choices of family firms. *Managerial Finance*, *40*(3), 254–275. https://doi.org/10.1108/MF-03-2013-0065

Gottardo, P., & Moisello, A. M. (2019). Family control and capital structure choices. In *Capital structure, earnings management, and risk of financial distress* (pp. 13–40). Springer. https://doi.org/10.1007/978-3-030-00344-9_2

Gugler, K. (2003). Corporate governance, dividend payout policy, and the interrelation between dividends, R&D, and capital investment. *Journal of Banking & Finance*, *27*, 1297–1321. https://doi.org/10.1016/S0378-4266(02)00258-3

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. N. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1–3), 389–422. https://doi.org/10.1023/A:1012487302797

Hauck, J., Suess-Reyes, J., Beck, S., Prügl, R., & Frank, H. (2016). Measuring socioemotional wealth in family-owned and-managed firms: A validation and short form of the FIBER scale. *Journal of Family Business Strategy*, *7*(3), 133–148. https://doi.org/10.1016/j.jfbs.2016.08.001

Huybrechts, J., Voordeckers, W., & Lybaert, N. (2013). Entrepreneurial risk taking of private family firms: The influence of a nonfamily CEO and the moderating effect of CEO tenure. *Family Business Review*, *26*(2), 161–179. https://doi.org/10.1177/0894486512469252

Jansen, K., Michiels, A., Voordeckers, W., & Steijvers, T. (2022). Financing decisions in private family firms: A family firm pecking order. *Small Business Economics*. https://doi.org/10.1007/s11187-022-00711-9

Kallmuenzer, A., Strobl, A., & Peters, M. (2018). Tweaking the entrepreneurial orientation–performance relationship in family firms: The effect of control mechanisms and family-related goals. *Review of Managerial Science*, *12*(4), 855–883. https://doi.org/10.1007/s11846-017-0231-6

Kets de Vries, M. F. K. (1993). The dynamics of family controlled firms: The good and the bad news. *Organizational Dynamics*, *21*(3), 59–71. https://doi.org/10.1016/0090-2616(93)90071-8

König, A., Kammerlander, N., & Enders, A. (2013). The family innovator's dilemma: How family influence affects the adoption of discontinuous technologies by incumbent firms. *Academy of Management Review*, *38*(3), 418–441. https://doi.org/10.5465/amr.2011.0162

Koropp, C., Kellermanns, F. W., Grichnik, D., & Stanley, L. (2014). Financial decision making in family firms: An adaptation of the theory of planned behavior. *Family Business Review*, *27*(4), 307–327. https://doi.org/10.1177/0894486514522483

Kotlar, J., Signori, A., De Massis, A., & Vismara, S. (2018). Financial wealth, socioemotional wealth, and IPO underpricing in family firms: A two-stage gamble model. *Academy of Management Journal*, *61*(3), 1073–1099. https://doi.org/10.5465/amj.2016.0256

Lardon, A., Deloof, M., & Jorissen, A. (2017). Outside CEOs, board control and the financing policy of small privately held family firms. *Journal of Family Business Strategy*, *8*(1), 29–41. https://doi.org/10.1016/j.jfbs.2017.01.002

Leitterstorf, M. P., & Rau, S. B. (2014). Socioemotional wealth and IPO underpricing of family firms. *Strategic Management Journal*, *35*(5), 751–760. https://doi.org/10.1002/smj.2236

Liu, Q., Luo, T., & Tian, G. G. (2015). Family control and corporate cash holdings: Evidence from China. *Journal of Corporate Finance*, *31*, 220–245. https://doi.org/10.1016/j.jcorpfin.2015.02.007

López-Delgado, P., & Diéguez-Soto, J. (2018). Indebtedness in family-managed firms: The moderating role of female directors on the board. *Review of Managerial Science*, *14*(4), 1–36. https://doi.org/10.1007/s11846-018-0307-y

López-Gracia, J., & Sánchez-Andújar, S. (2007). Financial structure of the family business: Evidence from a group of small Spanish firms. *Family Business Review*, *20*(4), 269–287. https://doi.org/10.1111/j.1741-6248.2007.00094.x

Martin, G., Campbell, J. T., & Gomez-Mejia, L. (2016). Family control, socioemotional wealth and earnings management in publicly traded firms. *Journal of Business Ethics*, *133*(3), 453–469. https://doi.org/10.1007/s10551-014-2403-5

Martin, G., & Gómez-Mejía, L. (2016). The relationship between socioemotional and financial wealth. Re-visiting family firm decision making. *Management Research*, *14*(3), 215–233. https://doi.org/10.1108/MRJIAM-02-2016-0638

Michiels, A., & Molly, V. (2017). Financing decisions in family businesses: A review of suggestions for developing the field. *Family Business Review*, *30*(4), 369–399. https://doi.org/10.1177/0894486517736958

Michiels, A., Uhlaner, L., & Dekker, J. (2017). The effect of family business professionalization on dividend payout. *Journal of Small Business and Enterprise Development*, *24*(4), 971–990. https://doi.org/10.1108/JSBED-01-2017-0023

Michiels, A., Voordeckers, W., Lybaert, N., & Steijvers, T. (2015). Dividends and family governance practices in private family firms. *Small Business Economics*, *44*(2), 299–314. https://doi.org/10.1007/s11187-014-9594-0

Miller, D., & le Breton–Miller, I. (2014). Deconstructing socioemotional wealth. *Entrepreneurship Theory and Practice*, *38*(4), 713–720. https://doi.org/10.1111/etap.12111

Mishra, C. S., & McConaughy, D. L. (1999). Founding family control and capital structure: The risk of loss of control and the aversion to debt. *Entrepreneurship Theory and Practice*, *23*(4), 53–64. https://doi.org/10.1177/104225879902300404

Molly, V., Uhlaner, L. M., De Massis, A., & Laveren, E. (2019). Family-centered goals, family board representation, and debt financing. *Small Business Economics*, *53*(1), 269–286. https://doi.org/10.1007/s11187-018-0058-9

Naldi, L., Cennamo, C., Corbetta, G., & Gomez-Mejia, L. (2013). Preserving socioemotional wealth in family firms: Asset or liability? The moderating role of business context. *Entrepreneurship Theory and Practice*, *37*(6), 1341–1360. https://doi.org/10.1111/etap.12069

Patel, P. C., & Chrisman, J. J. (2014). Risk abatement as a strategy for R&D investments in family firms. *Strategic Management Journal*, *35*(4), 617–627. https://doi.org/10.1002/smj.2119

Peláez-León, J. D., & Sánchez-Marín, G. (2022). Socioemotional wealth and human resource policies: Effects on family firm performance. *International Journal of Entrepreneurial Behavior & Research*, *28*(1), 109–135. https://doi.org/10.1108/IJEBR-05-2021-0404

Pindado, J., Requejo, I., & de la Torre, C. (2012). Do family firms use dividend policy as a governance mechanism? Evidence from the Euro zone. *Corporate Governance: An International Review*, *20*(5), 413–431. https://doi.org/10.1111/j.1467-8683.2012.00921.x

Poggi, J. M., & Tuleau, C. (2006). Classification supervisée en grande dimension. Application à l'agément de conduite automobile. *Revue de Statistique Appliquée*, *54*(4), 41–60.

Poutziouris, P. (2001). The views of family companies on venture capital: Empirical evidence from the UK small to medium-size enterprising economy. *Family Business Review*, *13*(3), 277–291. https://doi.org/10.1111/j.1741-6248.2001.00277.x

Prügl, R. (2019). Capturing the heterogeneity of family firms: Reviewing scales to directly measure socioemotional wealth. In *The Palgrave handbook of heterogeneity among family firms* (pp. 461–484). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-77676-7_17

Rakotomanonjy, A. (2003). Variable selection using SVM-based criteria. *Journal of Machine Learning Research*, *3*, 1357–1370.

Reina, W., Pla-Barber, J., & Villar, C. (2022). Socioemotional wealth in family business research: A systematic literature review on its definition, roles and dimensions. *European Management Journal*. https://doi.org/10.1016/j.emj.2022.10.009

Romano, C. A., Tanewski, G. A., & Smyrnios, K. X. (2001). Capital structure decision making: A model for family business. *Journal of Business Venturing*, *16*(3), 285–310. https://doi.org/10.1016/S0883-9026(99)00053-1

Sánchez-Marín, G., Meroño-Cerdán, Á. L., & Carrasco-Hernández, A. J. (2019). Formalized HR practices and firm performance: An empirical comparison of family and non-family firms. *The International Journal of Human Resource Management*, *30*(7), 1084–1110. https://doi.org/10.1080/09585192.2017.1289547

Schepers, J., Voordeckers, W., Steijvers, T., & Laveren, E. (2014). The entrepreneurial orientation–performance relationship in private family firms: The moderating role of socioemotional wealth. *Small Business Economics*, *43*(1), 39–55. https://doi.org/10.1007/s11187-013-9533-5

Schickinger, A., Bertschi-Michel, A., Leitterstorf, M. P., & Kammerlander, N. (2022). Same same, but different: Capital structures in single family offices compared with private equity firms. *Small Business Economics*, *58*(3), 1407–1425. https://doi.org/10.1007/s11187-021-00448-x

Schmid, T. (2013). Control considerations, creditor monitoring, and the capital structure of family firms. *Journal of Banking & Finance*, *37*(2), 257–272. https://doi.org/10.1016/j.jbankfin.2012.08.026

Schulze, W. S., & Kellermanns, F. W. (2015). Reifying socioemotional wealth. *Entrepreneurship, Theory and Practice*, *39*(3), 447–459. https://doi.org/10.1111/etap.12159

Schulze, W. S., Lubatkin, M. H., & Dino, R. N. (2003). Toward a theory of agency and altruism in family firms. *Journal of Business Venturing*, *18*(4), 473–490. https://doi.org/10.1016/S0883-9026(03)00054-5

Sciascia, S., Mazzola, P., & Kellermanns, F. W. (2014). Family management and profitability in private family-owned firms: Introducing generational stage and the socioemotional wealth perspective. *Journal of Family Business Strategy*, *5*(2), 131–137. https://doi.org/10.1016/j.jfbs.2014.03.001

Sciascia, S., Nordqvist, M., Mazzola, P., & De Massis, A. (2015). Family ownership and R&D intensity in small-and medium-sized firms. *Journal of Product Innovation Management*, *32*(3), 349–360.

Setia-Atmaja, L. (2010). Dividend and debt policies of family controlled firms. *International Journal of Managerial Finance*, *6*(2), 128–142. https://doi.org/10.1108/17439131011032059

Setia-Atmaja, L., Tanewski, G. A., & Skully, M. (2009). The role of dividends, debt and board structure in the governance of family controlled firms. *Journal of Business Finance & Accounting*, *36*(7–8), 863–898. https://doi.org/10.1111/j.1468-5957.2009.02151.x

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modelling. *Expert Systems with Applications*, *134*, 93–101. https://doi.org/10.1016/j.eswa.2019.05.028

Steijvers, T., & Niskanen, M. (2013). The determinants of cash holdings in private family firms. *Accounting & Finance*, *53*(2), 537–560. https://doi.org/10.1111/j.1467-629X.2012.00467.x

Stockmans, A., Lybaert, N., & Voordeckers, W. (2010). Socioemotional wealth and earnings management in private family firms. *Family Business Review*, *23*(3), 280–294. https://doi.org/10.1177/0894486510374457

Tabatabaee Malazi, H., & Davari, M. (2018). Combining emerging patterns with random forest for complex activity recognition in smart homes. *Applied Intelligence*, *48*(2), 315–330. https://doi.org/10.1007/s10489-017-0976-2

Uddin, M. S., Chi, G., Al Janabi, M. A. M., & Habib, T. (2020). Leveraging random forest in micro-enterprises credit risk modelling for accuracy and interpretability. *International Journal of Finance and Economics*, *1-17*, 3713–3729. https://doi.org/10.1002/ijfe.2346

Vandekerkhof, P., Steijvers, T., Hendriks, W., & Voordeckers, W. (2015). The effect of organizational characteristics on the appointment of nonfamily managers in private family firms:

The moderating role of socioemotional wealth. *Family Business Review*, *28*(2), 104–122. https://doi.org/10.1177/0894486513514274

Vandemaele, S., & Vancauteren, M. (2015). Nonfinancial goals, governance, and dividend payout in private family firms. *Journal of Small Business Management*, *53*(1), 166–182. https://doi.org/10.1111/jsbm.12063

Villalonga, B., Amit, R., Trujillo, M.-A., & Guzmán, A. (2015). Governance of family firms. *Annual Review of Financial Economics*, *7*(1), 635–654. https://doi.org/10.1146/annurev-financial-110613-034357

Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., & Zhang, B. (2018). Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Applied Water Science*, *8*(5), 1–12. https://doi.org/10.1007/s13201-018-0742-6

Wiseman, R. M., & Gomez-Mejia, L. R. (1998). A behavioral agency model of managerial risk taking. *Academy of Management Review*, *23*(1), 33–153. https://doi.org/10.5465/amr.1998.192967

Wu, Z., Chua, J. H., & Chrisman, J. J. (2007). Effects of family ownership and management on small business equity financing. *Journal of Business Venturing*, *22*(6), 875–895. https://doi.org/10.1016/j.jbusvent.2006.07.002

Zellweger, T. M., Kellermanns, F. W., Chrisman, J. J., & Chua, J. H. (2012). Family control and family firm valuation by family CEOs: The importance of intentions for transgenerational control. *Organization Science*, *23*(3), 851–868. https://doi.org/10.1287/orsc.1110.0665

AUTHOR BIOGRAPHIES

Susana Álvarez-Díez is a Full Professor at the Department of Quantitative Methods for the Economy and Business at the University of Murcia in Spain. He received her PhD in Economics from the University of Alicante. Her main research areas include machine learning classification models and machine learning regression models.

J. Samuel Baixauli-Soler is a Full Professor at the Department of Management and Finance at the University of Murcia in Spain. He received his PhD in Business Administration from the University of Valencia. His main research areas include computational finance and business finance.

María Belda-Ruiz is a Full Professor at the Department of Management and Finance at the University of Murcia in Spain. She received her International PhD in Business Science from the University of Murcia. Her main research areas include financial economics, corporate governance, and financial decisions in family firms.

Gregorio Sánchez-Marín is a Full Professor at the Department of Economics and Business at the University of Alcala in Spain. He received his PhD in Business Administration from the University of Murcia. His main research areas include executive compensation, SMEs internationalization processes, and human resource management in family firms.

How to cite this article: Álvarez-Díez, S., Baixauli-Soler, J. S., Belda-Ruiz, M., & Sánchez-Marín, G. (2023). Variable selection for classification and forecasting of the family firm's socioemotional wealth. *Journal of Forecasting*, *42*(8), 2063–2078. https://doi.org/10.1002/for.3008